JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **15**: 59–81 (2000)

ASYMPTOTICALLY PERFECT AND RELATIVE CONVERGENCE OF PRODUCTIVITY

BART HOBIJN[a]* AND PHILIP HANS FRANSES[b]
[a]*Department of Economics, New York University, 269 Mercer Street, New York, NY 10003, USA*
[b]*Econometric Institute, Erasmus University Rotterdam, PO Box 1738, 3000 DR, Rotterdam, The Netherlands*

SUMMARY

In this paper we examine the extent to which countries are converging in per capita productivity levels. We propose to use cluster analysis in order to allow for the endogenous selection of converging countries. We formally define convergence in a time series analytical context, derive the necessary and sufficient conditions for convergence, and introduce a cluster analytical procedure that distinguishes several convergence clubs by testing for these conditions using a multivariate test for stationarity. We find a large number of relatively small convergence clubs, which suggests that convergence might not be such a widespread phenomenon.

1. INTRODUCTION

Despite the availability of many complementary and competing definitions of convergence, economists seem to agree that no matter what definition of convergence is used, unconditional convergence of per capita productivity levels for all the economies in the world does not exist. This led Baumol (1986) to suggest that there might be several groups of countries, known as convergence clubs, within which we observe convergence, but that do not converge to each other. If these convergence clubs really exist, then there are basically two main issues. The first is why some countries grow faster than others, while the second has to do with the number of convergence clubs and how many countries they contain.

The question why some countries grow faster than others has been the focus of the extensive empirical literature on conditional convergence. Mankiw, Romer and Weil (1992), Barro and Sala-i-Martin (1992), Islam (1995), Canova and Marcet (1995), and Caselli *et al*. (1996), for example, all find evidence that countries are converging to their individual balanced growth paths, which depend on, among others, human capital levels, political stability and inequality, the savings rate and the population growth rate. See also Durlauf and Quah (1998) for a recent survey of the existing empirical literature on economic growth.

The second question of how many convergence clubs there are and how many countries they contain did not get as much attention. This is probably due to the fact that an answer to this question requires the application of non-standard econometric techniques that allow us to divide the whole sample of countries into smaller clubs. In this paper we introduce such a technique and apply it to data on real GDP per capita from the Penn World Table, described in Summers and

* Correspondence to: Bart Hobijn, New York University, Department of Economics, 269 Mercer Street, 7th Floor, New York, NY 10003, USA; e-mail: bart.hobijn@econ.nyu.edu

Contract/grant sponsor: C. V. Starr Center, New York University.

Received 16 June 1997
Revised 1 February 1999

Heston (1991), and to data used by Bernard and Durlauf (1995). The technique that we introduce is a cluster algorithm that allows for the endogenous selection of converging countries.

Our approach of dividing a large sample of countries into subgroups is certainly not new. Durlauf and Johnson (1995) continue in the tradition of the conditional convergence literature and use a regression-tree procedure to determine threshold levels of initial per capita income and literacy rates that imply groups of countries that satisfy common linear cross-sectional regression equations. However, instead of linear cross-section regression equations, we use time-series analytical techniques to cluster countries on the basis of common asymptotic behaviour of their per capita income levels.

In particular, we will focus on three types of common asymptotic behaviour of per capita income levels. The first, introduced by Bernard and Durlauf (1995), is known as asymptotically perfect convergence and implies that, independently of their current income levels, we might expect countries to converge to identical log per capita income levels. The second type, which amounts to an extension of Bernard and Durlauf (1995) and which we will call asymptotically relative convergence, implies that, independently of the current situation, we might expect the difference between countries' log per capita income levels to converge to a finite constant. These concepts are closely related to the dynamics of the distribution of log per capita income levels across countries, as studied by, for example, Quah (1996a,b), Bianchi (1997), Jones (1997), and Pritchett (1997). We will show that, under certain assumptions, whenever a group of countries exhibits neither asymptotically perfect nor asymptotically relative convergence, then the variance of their cross-sectional distribution of log per capita income levels will become unbounded. Third, and finally, we briefly consider convergence of growth rates.

Our cluster algorithm has three main properties. The first is

Property 1 *Testing for necessary and sufficient conditions*. The cluster algorithm tests for the necessary and sufficient conditions that have to be satisfied in order for our definitions of convergence to hold.

This means that, when our inference is correct, we indeed find convergence clubs and not clubs of countries that might possibly converge because they satisfy only necessary conditions for convergence. As Quah (1993) pointed out, the latter is the case for many of the cross-sectional studies of convergence.

Property 2 *Cluster algorithm is consistent*. When the number of observations per series goes to infinity, we will find the true underlying convergence clubs with certainty.

Combined with Property 1, this basically implies that we want our cluster algorithm to find the true convergence clubs whenever we have enough data. This is very much equivalent to the standard concept of consistency of point estimates. Finally,

Property 3 *Independence of ordering of series*. The cluster algorithm does not depend on the ordering of the series.

The reason that we impose this third property is immediately obvious. We do not want our results to depend on the arbitrary ordering of our data.

The results obtained with our cluster algorithm indicate that convergence might not be a widespread phenomenon. Both for the data from the Penn World Table, which cover 112 countries for the period 1960–1989, and the data used by Bernard and Durlauf (1995), which cover 15 OECD countries over the period 1900–1987, we find a large number of rather small convergence clubs.

The structure of the paper is as follows. In Section 2 we introduce our three convergence concepts, i.e. asymptotically perfect convergence, asymptotically relative convergence, and convergence in growth rates. We derive the necessary and sufficient conditions for these types of convergence and illustrate what these definitions imply for the asymptotic behaviour of the cross-sectional distribution of countries' log per capita income levels. In Section 3 we introduce a multivariate generalization of the test for the null hypothesis of stationarity of a time series introduced by Kwiatkowski *et al.* (1992, KPSS hereafter). This test can be used to test for the null hypothesis that a group of countries is converging. We derive its asymptotic distribution and show that it is consistent under the relevant alternatives. In Section 4 we introduce our cluster algorithm and show that it satisfies the properties listed above. We present our empirical results in Section 5 and conclude with some remarks and suggestions for further research in Section 6. Two appendices follow. The first contains the proofs of the propositions and the second presents a complete list of the convergence clubs found.

2. DEFINITIONS OF CONVERGENCE

There are two basic approaches to the empirical analysis of the convergence hypothesis, as described in Bernard and Durlauf (1996). On the one hand, there are cross-section analyses that focus on a negative correlation between initial productivity levels and growth rates. On the other hand, there are time series analytical approaches that focus mainly on the asymptotic properties of the disparities in productivity levels between economies. Clearly, for a cross-section analysis we need data for a sufficiently large number of countries in order to be able to make reliable inference about the existence or non-existence of convergence in the sample studied. Since we want to allow for the possible finding of arbitrarily small convergence clubs, with as the most extreme case a single country that does not converge to any other, we cannot use this cross-section approach. Hence, we consider the data in the time dimension, and have to base our analysis on the asymptotic properties of the productivity disparities between the various countries in our data set.

Throughout this paper we will consider the log per capita income levels of a sample of n countries for T years of data. Let y_{it} be the log per capita income level of country i in period t. We assume that the multivariate process $\mathbf{y}_t = [y_{1t}, \ldots, y_{nt}]' \in \mathbb{R}^n$ has the following representation:

$$\underset{n\times 1}{\mathbf{y}_t} = \underset{n\times 1}{\mathbf{a}} + \underset{n\times 1}{\mathbf{b}}\, t + \underset{n\times m^*}{\mathbf{D}^*} \sum_{s=0}^{t-1} \underset{m^*\times 1}{\mathbf{v}_s^*} + \underset{n\times 1}{\mathbf{u}_t^*} \tag{1}$$

where $\mathbf{v}_t^*$ represents the first difference of the $m^* \in \{0, \ldots, n\}$ common trends in $\mathbf{y}_t$.

In this case, each univariate process y_{it} equals

$$y_{it} = a_i + b_i t + \sum_{l=1}^{m^*} D_{il}^* \left(\sum_{s=0}^{t-1} v_{is}^* \right) + u_{it}^* \tag{2}$$

where a_i, b_i, v_{it}^* and u_{it}^* are the ith elements of the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{v}_t^*$ and $\mathbf{u}_t^*$ respectively, while D_{il}^* is the (i, l)th element of $\mathbf{D}^*$.

The three types of convergence that we will consider are all based on the properties of the disparities of the per capita productivity levels between two countries, say i and j. We will denote this disparity by $d_{(i,j)t} = y_{it} - y_{jt}$. The first definition of convergence that we consider is that of asymptotically perfect convergence, which reads:

Definition 1 *Asymptotically perfect convergence*. Countries i and j are assumed to converge asymptotically perfectly, whenever $d_{(i,j)t}$ is zero mean stationary.

This definition is identical to that of convergence in Bernard and Durlauf (1995) and it implies that at every moment in time we expect the log of the per capita income levels to ultimately become equal, no matter what are the current and past levels of per capita income of the countries considered. Baumol *et al.* (1994), however, has argued that this is too strong a condition, because we are more likely to find that the process of convergence is limited and that, once countries are relatively close to each other, the process of convergence will stop. A theoretical model that predicts this is that of Barro and Xala-i-Martin (1995) in which imitating countries catch up with countries that innovate. However, costs of imitation limit the process of convergence and they will not catch up completely. This argument suggests that it would be worth while to look at a definition of convergence that implies that, at every moment in time and independently of the current and past per capita income levels, we might expect countries to ultimately converge to constant relative per capita productivity levels. We therefore complement Bernard and Durlauf's (1995) definition of convergence with

Definition 2 *Asymptotically relative convergence*. Countries i and j are assumed to converge asymptotically relatively whenever $d_{(i,j)t}$ is level stationary.

Finally, we consider an even less strong definition of convergence, namely that at every moment in time and independently of the current and past per capita income levels, we might expect countries to ultimately converge to the same growth rates of per capita productivity. That is:

Definition 3 *Convergence of growth rates*. Countries i and j are assumed to converge in growth rates whenever $\Delta d_{(i,j)t}$ is zero mean stationary, where Δ denotes the first differencing operator.

Given equation (1), asymptotically perfect convergence implies asymptotically relative convergence which, in turn, implies convergence in growth rates. This can be easily seen by considering the implied parameter restrictions in equation (2). These parameter restrictions are listed in Table I.

It is illustrative to see what these definitions and restrictions imply for the variance of the cross-country per capita income distribution. For a subsample of n^* countries we are interested in the expected long-run cross-sectional variance of their log per capita income levels, which is

Table I. Parameter restrictions in equation (2) for various types of convergence

Type of convergence	Parameter restrictions		
Asymptotically perfect	$a_i = a_j$,	$D^*_{il} = D^*_{jl}$ for all $l = 1, \ldots, m^*$,	$b_i = b_j$
Asymptotically relative		$D^*_{il} = D^*_{jl}$ for all $l = 1, \ldots, m^*$,	$b_i = b_j$
Growth rates			$b_i = b_j$

defined as

$$V = \lim_{T\to\infty} E_t\left[\frac{1}{n^*}\sum_{i=1}^{n^*}(y_{i,t+T} - \bar{y}_{t+T})^2\right]$$

where

$$\bar{y}_t = \frac{1}{n^*}\sum_{i=1}^{n^*} y_{it}$$

It can be easily seen that this limit goes to infinity whenever this subsample contains two countries i and j for which either $D^*_{il} \neq D^*_{jl}$ for some l or $b_i \neq b_j$. Basically, this implies that the cross-sectional variance of log per capita income levels will go to infinity whenever these levels are subject to different stochastic or deterministic trends. Jones (1997) already argued that convergence of growth rates is a necessary condition for the world income distribution to settle down to a stable non-degenerate distribution. An important implication of the above result is that convergence of growth rates is only a necessary but not a sufficient condition because it does not take into account the possible existence of common stochastic trends.

Contrary to convergence of growth rates, asymptotically perfect and relative convergence do imply restrictions on the stochastic trends underlying log per capita income levels. In particular, they imply that two countries, say i and j, can only converge whenever the stochastic trends that drive their log per capita income levels are cointegrated with cointegrating vector $[1\ \ -1]$, which is implied by the restriction $D^*_{il} = D^*_{jl}$ for all $l = 1, \ldots, m^*$. As a consequence, we obtain that if the subsample of n^* countries converge asymptotically relatively, the variance

$$V = \frac{1}{n^*}\sum_{i=1}^{n^*}(a_i - \bar{a})^2 + E\left[\frac{1}{n^*}\sum_{i=1}^{n^*}(u^*_{it} - \bar{u}^*_t)^2\right]$$

where

$$\bar{a} = \frac{1}{n^*}\sum_{i=1}^{n^*} a_i$$

and

$$\bar{u}^*_t = \frac{1}{n^*}\sum_{i=1}^{n^*} u^*_{it}$$

is clearly finite whenever n^* is finite. Similarly, whenever the subsample of countries converge asymptotically perfectly, this variance

$$V = E\left[\frac{1}{n^*}\sum_{i=1}^{n^*}(u^*_{it} - \bar{u}^*_t)^2\right]$$

is again finite. Hence, contrary to convergence of growth rates, asymptotically perfect and relative convergence imply that the cross-country income distribution settles down to a stable non-degenerate distribution. This is the reason why, in the remainder of the paper, we will focus on asymptotically perfect and relative convergence rather than on convergence of growth rates.

Since our aim is to use a multivariate test that enables us to test for the joint null hypothesis that a group of countries is converging, we will proceed by rewriting the parameter restrictions of Table I in terms of vector processes. Consider again a subsample of n^* countries and the associated vector process of log per capita income levels, i.e. $\mathbf{y}_t^* = [y_{1t}, \ldots, y_{n^*t}]' \in \mathbb{R}^{n^*}$. Defining

$$\underset{(n^*-1)\times n^*}{\mathbf{M}_{n^*}} = \begin{bmatrix} 1 & -1 & 0 & \ldots & \ldots & 0 \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & 1 & -1 \end{bmatrix}$$

and $\mathbf{x}_t \equiv \mathbf{M}_{n^*}\mathbf{y}_t^* \in \mathbb{R}^{n^*-1}$, we obtain that the parameter restrictions listed in Table I imply

Hypothesis 1 *Zero mean stationarity*. If the n^* countries considered are converging asymptotically perfectly, then $\mathbf{x}_t$ is zero mean stationary.

Hypothesis 2 *Level stationarity*. If the n^* countries considered are converging asymptotically relatively, then $\mathbf{x}_t$ is level stationary.

Hence, we will thus have to consider a multivariate test for stationarity. In the following section we will introduce this test, derive its asymptotic distribution and show that the derived test statistic is invariant to the ordering of the data. The latter is important, because this result implies that when we apply this test in our cluster algorithm, the algorithm will satisfy property 3 mentioned in the Introduction. One of the asymptotic properties of the test introduced below is that it is consistent. Consistency of the test means that if the n^* countries considered do not converge, then the null hypothesis of zero mean stationarity or level stationarity—whichever is relevant—is rejected with certainty as the sample size, T, goes to infinity. Hence, as the sample size becomes large enough, we will not find any spurious convergence relationships.

3. MULTIVARIATE STATIONARITY TEST

The test that we introduce in this section will be applied to $\mathbf{x}_t \equiv \mathbf{M}_{n^*}\mathbf{y}_t^* \in \mathbb{R}^k$, where $k = n^* - 1$. This implies that $\mathbf{x}_t$ has a representation similar to that of $\mathbf{y}_t$, in the sense that

$$\underset{k\times 1}{\mathbf{x}_t} = \underset{k\times 1}{\boldsymbol{\alpha}} + \underset{k\times 1}{\boldsymbol{\beta}}\, t + \underset{k\times m}{\mathbf{D}} \sum_{s=0}^{t-1} \underset{m\times 1}{\mathbf{v}_s} + \underset{k\times 1}{\mathbf{u}_t}$$

where the accumulated $\mathbf{v}_s$ constitute the $m \in \{0, \ldots, k\}$ common trends in $\mathbf{x}_t$ and the matrix $\mathbf{D}$ has full column rank. The process $\boldsymbol{\eta}_t = [\mathbf{u}_t'\mathbf{v}_t']'$ is covariance stationary and we will assume that it has the following $MA(\infty)$-representation:

$$\underset{(k+m)\times 1}{\boldsymbol{\eta}_t} = \sum_{s=0}^{\infty} \underset{(k+m)\times(k+m)}{\boldsymbol{\Psi}_s} \underset{(k+m)\times 1}{\boldsymbol{\varepsilon}_{t-s}} = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_s$$

where $\{s \cdot \boldsymbol{\Psi}_s\}_{s=0}^{\infty}$ is absolutely summable, as defined in Hamilton (1994, p. 547). $\boldsymbol{\varepsilon}_t$ is an i.i.d. sequence with mean zero, $E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'] = \boldsymbol{\Omega}$, with Choleski factorization $\boldsymbol{\Omega} = \mathbf{PP}'$. Additionally, we will assume that $\boldsymbol{\varepsilon}_t$ has finite fourth moments, such that the functional central limit theorem applies. Furthermore, we define the matrices

$$\boldsymbol{\Lambda} = \boldsymbol{\Psi}(1)\mathbf{P}$$

and

$$\mathbf{G} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$$

Our definitions of convergence imply the restrictions on the parameters of the process $\mathbf{x}_t$ listed in Table II. Here, again, zero mean stationarity corresponds to asymptotically perfect convergence, while level stationarity corresponds to asymptotically relative convergence. In the rest of this section we will introduce two multivariate stationarity tests that can be used to test for these respective null hypotheses. These tests are multivariate generalizations of the stationarity test introduced by KPSS.

Table II. Null hypotheses

	Hypothesis	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	m
H_0	Zero mean stationarity	$\mathbf{0}$	$\mathbf{0}$	0
H_μ	Level stationarity	Unrestricted	$\mathbf{0}$	0

Definition 4 Define $\mathbf{e}_t = \mathbf{x}_t - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\beta}}t$, where $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are obtained from a regression of $\mathbf{x}_t$ on an intercept and a deterministic trend. Define $\hat{\mathbf{G}}_l$ as the consistent Newey–West (1987) estimator of the first k rows and columns of $\mathbf{G}$, under the null hypotheses. That is,

$$\hat{\mathbf{G}}_l = \hat{\boldsymbol{\Gamma}}_0 + \sum_{s=1}^{l}\left(1 - \frac{s}{l+1}\right)(\hat{\boldsymbol{\Gamma}}_s + \hat{\boldsymbol{\Gamma}}_s')$$

where

$$\hat{\boldsymbol{\Gamma}}_s = T^{-1}\sum_{t=s+1}^{T} \mathbf{e}_t\mathbf{e}_{t-s}'$$

and the bandwidth parameter l is assumed to satisfy that $T^{-1/2}l \to 0$ as $T \to \infty$.

Definition 5 *Test for zero mean stationarity*. Suppose $\{\mathbf{x}_t\}_{t=1}^{T}$ is as defined above and define the partial sum process

$$\mathbf{S}_t \equiv \sum_{s=1}^{t} \mathbf{x}_s$$

such that $\mathbf{S}_t \in \mathbb{R}^k$, then we will consider the following test statistic for zero mean stationarity:

$$\bar{\varpi}_0 = T^{-2} \sum_{t=1}^{T} \mathbf{S}_t' [\hat{\mathbf{G}}_l]^{-1} \mathbf{S}_t$$

Definition 6 *Test for level stationarity*. Suppose $\{\mathbf{x}_t\}_{t=1}^T$ is as defined above and define

$$\mathbf{z}_t \equiv \mathbf{x}_t - \left(\frac{1}{T} \sum_{s=1}^{T} \mathbf{x}_s \right)$$

and

$$\tilde{\mathbf{S}}_t = \sum_{s=1}^{t} \mathbf{z}_s$$

such that $\tilde{\mathbf{S}}_t \in \mathbb{R}^k$, then we will consider the following test statistic for level stationarity:

$$\bar{\varpi}_\mu = T^{-2} \sum_{t=1}^{T} \tilde{\mathbf{S}}_t' [\hat{\mathbf{G}}_l]^{-1} \tilde{\mathbf{S}}_t$$

In order to prove that our cluster algorithm satisfies property 3 stated in the Introduction, namely the independence of the ordering of the series, we prove the following result in Appendix A:

Proposition 1 *Invariance of test statistics to non-singular linear transformations*. Consider the vector process $\mathbf{x}_t \in \mathbb{R}^k$ and the linear transformation

$$\mathbf{x}_t^* \equiv \underset{k \times k}{\mathbf{A}} \mathbf{x}_t \in \mathbb{R}^k$$

where $\mathbf{A}$ is non-singular. Denote the stationarity tests obtained using $\mathbf{x}_t$ as $\bar{\varpi}_0$ and $\bar{\varpi}_\mu$ respectively, and those obtained with $\mathbf{x}_t^*$ as $\bar{\varpi}_0^*$ and $\bar{\varpi}_\mu^*$, then $\bar{\varpi}_0 = \bar{\varpi}_0^*$ and $\bar{\varpi}_\mu = \bar{\varpi}_\mu^*$.

Let $\mathbf{W}(r) : r \in [0, 1] \to \mathbb{R}^k$ be a k-dimensional standard Brownian motion,[1] and $\mathbf{V}(r) : r \in [0, 1] \to \mathbb{R}^k$ be a first-level k-dimensional standard Brownian bridge, such that $\mathbf{V}(r) \equiv \mathbf{W}(r) - r\mathbf{W}(1)$, then we obtain the following asymptotic distributions of $\bar{\varpi}_0$ and $\bar{\varpi}_\mu$ under their respective null hypotheses:

Proposition 2 *Asymptotic distribution of* $\bar{\varpi}_0$. Under H_0, the test statistic $\bar{\varpi}_0$ has the following asymptotic distribution:

$$\bar{\varpi}_0 \to \int_0^1 \mathbf{W}(r)' \mathbf{W}(r) \, \mathrm{d}r$$

where $\to$ denotes weak convergence.

Proposition 3 *Asymptotic distribution of* $\bar{\varpi}_\mu$. Under H_μ, the test statistic $\bar{\varpi}_\mu$ has the following asymptotic distribution:

$$\bar{\varpi}_\mu \to \int_0^1 \mathbf{V}(r)' \mathbf{V}(r) \, \mathrm{d}r$$

[1] For a definition of $\mathbf{W}(r)$ see, for example, Hamilton (1994, p. 544).

Table III. Percentiles of asymptotic distributions

	$\bar{\varpi}_0$				$\bar{\varpi}_\mu$			
k	0·50	0·90	0·95	0·99	0·50	0·90	0·95	0·99
1	0·296	1·118	1·650	2·773	0·117	0·341	0·462	0·727
2	0·763	2·065	2·626	3·893	0·275	0·612	0·753	1·078
3	1·256	2·817	3·428	4·832	0·440	0·843	1·002	1·370
4	1·749	3·522	4·187	5·762	0·601	1·062	1·244	1·626
5	2·255	4·217	4·949	6·602	0·769	1·278	1·467	1·874
10	4·691	7·434	8·392	10·286	1·598	2·285	2·514	3·022
20	9·699	13·429	14·585	17·079	3·251	4·215	4·528	5·186
30	14·712	19·132	20·640	23·348	4·919	6·054	6·429	7·187
40	19·649	24·799	26·397	29·803	6·584	7·891	8·283	9·179
50	24·643	30·446	32·175	35·680	8·236	9·703	10·134	11·018

Values based on Monte Carlo simulations with sample size 10,000 and 100,000 replications.

The proofs of these propositions are straightforward multivariate generalizations of those in KPSS. These distributions differ for different k and their 50th, 90th, 95th and 99th percentiles are listed for various values of k in Table III.

The asymptotic behaviour of the test statistics under the alternative hypotheses is derived in the following proposition, which is proved in Appendix A.

Proposition 4 *Consistency of* $\bar{\varpi}_0$ *and* $\bar{\varpi}_\mu$. If $\boldsymbol{\alpha} \neq \mathbf{0}$ and/or $\boldsymbol{\beta} \neq \mathbf{0}$ and/or $m \neq 0$, then $\bar{\varpi}_0 \to \infty$ as $T \to \infty$. Similarly, if $\boldsymbol{\beta} \neq \mathbf{0}$ and/or $m \neq 0$, then $\bar{\varpi}_\mu \to \infty$ as $T \to \infty$. In particular, the rate of consistency is always at least T/l. Here l is the bandwidth of the Bartlett window used to estimate $\hat{\mathbf{G}}_l$.

This result is more general than the consistency result in KPSS, because it shows consistency not only under the alternative of a stochastic trend, when $m \neq 0$, but also under the additional alternatives of a deterministic trend and a non-zero intercept. This extension is important, since it means that, as the sample size becomes arbitrarily large, the test statistic will reject a wrong null hypothesis with certainty under all alternatives we consider.

According to the results of Propositions 1 through 4, the test statistics introduced in this section thus have the following three properties: (1) they test for the *null hypothesis* of convergence; (2) if a group of countries does not converge, then the tests will reject the null hypothesis with certainty as the number of observations per country goes to infinity; (3) the test is invariant to any non-singular linear transformation of the variables. In particular, this implies that it is also invariant to the ordering of the variables in $\mathbf{y}_t^*$. For these three reasons these tests seem a perfect basis for the cluster algorithm that we will introduce in the next section.

4. CLUSTER ALGORITHM

In this section we will introduce our cluster algorithm. After its introduction, we illustrate why it satisfies the properties stated in the Introduction and, finally, we comment on some of its limitations when applied to relatively short time series. Since asymptotically perfect convergence implies asymptotically relative convergence, our cluster algorithm consists of two parts. In the first part we cluster the countries in our sample on the basis of zero mean stationarity, then we cluster these asymptotically perfect convergence clubs by testing for level stationarity, yielding asymptotically relative convergence clubs. The cluster algorithm that we will use consists of the following steps:

Cluster algorithm We will denote the set of countries in cluster i by $k(i)$, the number of countries in cluster i by k_i, the number of countries in our sample by n, and the vector process containing the log of real GDP per capita of the countries in cluster i by $\mathbf{y}_t^{(i)} \in \mathbb{R}^{k_i}$. Furthermore, let $\bar{\varpi}_0^{(i,j)}$ and $\bar{\varpi}_\mu^{(i,j)}$ be the tests for zero mean stationarity and level stationarity applied to $\mathbf{x}_t^{(i,j)} \equiv \mathbf{M}_{k_i+k_j}[\mathbf{y}_t^{(i)\prime}\mathbf{y}_t^{(j)\prime}]' \in \mathbb{R}^{k_i+k_j-1}$, i.e. $\bar{\varpi}_0^{(i,j)}$ and $\bar{\varpi}_\mu^{(i,j)}$ are the tests for asymptotically perfect and relative convergence of clusters i and j. Finally, let $p_0^{(i,j)}$ and $p_\mu^{(i,j)}$ denote the p-values — excess probabilities — of $\bar{\varpi}_0^{(i,j)}$ and $\bar{\varpi}_\mu^{(i,j)}$ calculated from their asymptotic distributions which in turn depend on $k_i + k_j$. The cluster algorithm requires the choice of a critical p-value, loosely interpretable as a 'significance level', denoted by $p_{\min} \in (0, 1)$. Using these definitions, the countries can be clustered using the following steps:

(1) Initialize the cluster algorithm by setting $k(i) = \{i\}$ for all $i = 1, \ldots, n$.
(2) If there is no combination of i and j for which $p_0^{(i,j)} > p_{\min}$, i.e. for all pairs of clusters we reject asymptotically perfect convergence, then we stop and go to step (5). If there is such a combination, then we proceed with step (3).
(3) Choose i and j, such that $i < j$ and i and j correspond to the largest $p_0^{(i,j)}$. These two clusters are assumed to be 'most likely' to converge.
(4) Combine the clusters i and j by redefining $k(i)$ to equal $k(i) \cup k(j)$ and discarding cluster j by setting $k(j) = \emptyset$. Return to step (2).
(5) The clusters obtained are asymptotically perfectly convergence clubs.
(6) Proceed with clustering countries using the criterion of level stationarity rather than that of zero mean stationarity: Again, cluster using steps (2) through (4), but now use $p_\mu^{(i,j)}$ instead of $p_0^{(i,j)}$.
(7) The clusters obtained in step (6) are asymptotically relative convergence clubs.

The first part of the algorithm, in which we obtain the asymptotically perfectly convergence clubs, consists of steps (1) through (5), while the second part consists of steps (6) and (7).

Since the test statistics $\bar{\varpi}_0^{(i,j)}$ and $\bar{\varpi}_\mu^{(i,j)}$ are used in the cluster algorithm to construct asymptotically perfect and relative convergence clubs respectively and these tests test for the necessary and sufficient conditions for convergence, it is immediately obvious that the algorithm satisfies Property 1. It clusters on the basis of the necessary and sufficient conditions for convergence. Since $\bar{\varpi}_0^{(i,j)}$ and $\bar{\varpi}_\mu^{(i,j)}$ are invariant to the ordering of the series, so is the cluster algorithm and it thus satisfies Property 3. That the algorithm satisfies Property 2, i.e. consistency, is less obvious. Consistency of the algorithm can be achieved by choosing the critical p-value, i.e. $p_{\min}$, as a function of the sample size such that $p_{\min}(T) \to 0$ as $T \to \infty$. Such a choice will ensure that, asymptotically, the test never rejects a true null hypothesis of convergence. Additionally, when we denote the critical values for the tests with dimension k associated with this 'significance level' as $c_0(T, k)$ and $c_\mu(T, k)$, then, in order for the test to maintain its consistency, $p_{\min}(T)$ also has to satisfy that $c_0(T, k)l/T \to 0$ and $c_\mu(T, k)l/T \to 0$ for all $k \in \{1, \ldots, n-1\}$. That is, the rate at which these critical values go to infinity does not exceed the rate of consistency of the test statistics.

Hence, independently of the ordering of our series, the cluster algorithm introduced above will find the true convergence clubs with certainty when the sample size T goes to infinity. In practice, however, we deal with relatively small samples and it is thus interesting to consider some of the small sample limitations of the procedure.

Application of the procedure implies the choice of two parameters: the critical p-value $p_{\min}$, and the bandwidth parameter l. The effect of the choice of $p_{\min}$ is that the smaller this p-value is chosen, the less likely we are to reject the null hypothesis of convergence and the larger are the convergence clubs that we find. The bandwidth choice l has no obvious effect on the size of the convergence clubs found. It affects $\hat{\mathbf{G}}_l$ and therefore the value of the test statistic for all combinations of countries and clusters. Monte Carlo results for the univariate version of the KPSS test, as presented by KPSS and Hobijn, Franses, and Ooms (1998), suggest that, in small samples, the size of the test depends significantly on the choice of l. In the case of a small sample, robustness of the results of the cluster algorithm for various choices of l is thus something that needs to be checked.

There is another important limitation of the procedure in the case of small samples. If $T < n$, as is the case for the data from the Penn World Table to which we apply our algorithm below, the cluster algorithm could find combinations of clusters, say i and j, for which $k_i + k_j - 1 \geqslant T$. In that case $\hat{\mathbf{G}}_l$ will be singular, because no (multivariate) inference about such a large cluster is possible on the basis of the available data. The cluster algorithm can simply not handle such a case. In general, one has to bear in mind that the larger the cluster, the less reliable the inference.

5. RESULTS

In this section we will apply our cluster algorithm to two data sets. The first is the Penn World Table, Mark 5.5, as described in Summers and Heston (1991), which consists of data for 112 countries and for the period 1960–1989. The second, used by Bernard and Durlauf (1995), consists of longer time series covering the period 1900–1989, but contains data only for 15 industrialized countries. The empirical results presented in this section consist of two parts. First, we consider the robustness of our results for different choices of the bandwidth parameter l. Second, we present the convergence clubs that we obtain and compare our results with other empirical studies and theories on convergence.

5.1. Sensitivity Analysis

As noted above, when we are dealing with short time series the outcome of the cluster procedure is likely to differ for various choices of the bandwidth parameter l. Basically, this implies that in small samples it is hard to obtain a reliable estimate of the long-run variance $\mathbf{G}$ needed for the calculation of our test statistics. The outcome of the cluster algorithm can vary in two ways. First, the degree of convergence, i.e. the number of convergence clubs found, can change. Second, the composition of the convergence clubs can change.

In order to be able to compare the compositions of the convergence clubs for different choices of the bandwidth parameter, l, we introduce a descriptive statistic, $r_{a,b}$, which we will call a cluster correlation. The cluster correlation measures the degree of overlap of two outcomes, a and b, of

Table IV. Sensitivity of results for the Penn World Table

	Bandwidth (l)		
	1	2	3
# clubs	68 \ 42	63 \ 42	61 \ 39
l		Cluster correlations	
1		0·6515	0·5632
2	0·5287		0·6484
3	0·4753	0·5986	

All results are for $p_{\min} = 0{\cdot}01$.
clubs:
perfectly converging clubs; \# relatively converging clubs.
Cluster correlations:
Above diagonal: cluster correlations for perfectly converging clubs.
Below diagonal: cluster correlations for relatively converging clubs.

Table V. Sensitivity of results of the Bernard and Durlauf data

	Bandwidth (l)					
	1	2	3	4	5	6
# clubs	12 \ 8	10 \ 7	9 \ 7	9 \ 7	9 \ 7	9 \ 6
l			Cluster correlations			
1		0·6866	0·6606	0·6606	0·6606	0·4671
2	0·6687		0·9622	0·9622	0·9622	0·8784
3	0·5655	0·8734		1·0000	1·0000	0·9258
4	0·5655	0·8734	1·0000		1·0000	0·9258
5	0·5655	0·8734	1·0000	1·0000		0·9258
6	0·5857	0·7833	0·8969	0·8969	0·8969	

All results are for $p_{\min} = 0{\cdot}01$.

the cluster algorithm for the same set of countries and is defined by

$$r_{a,b} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j\neq i}^{n}\delta_{i,j}^{a}\delta_{i,j}^{b}}{\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n}\delta_{i,j}^{a}\right)^{1/2}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n}\delta_{i,j}^{b}\right)^{1/2}}}$$

where $\delta_{i,j}^{x}$ equals one if countries i and j are in the same convergence club for outcome x and zero otherwise.[2]

Tables IV and V contain the results of our sensitivity analysis. Since all the results we present in this section suggest relatively little convergence, we present them for $p_{\min} = 0{\cdot}01$; choosing a higher $p_{\min}$ will lead to finding even less convergence. As can be seen from Table IV the number of convergence clubs found for the Penn World Table data is not very sensitive to the choice of the bandwidth l. This suggests that the degree of convergence found is relatively robust to the

[2] The cluster correlation is not defined in case one of the outcomes a or b is that every country forms its own cluster. Since this case implies that there will be no convergence relationships, in this case a and b can have no convergence relationships in common and $r_{a,b} = 0$ by definition.

bandwidth. This is much less the case for the composition of the convergence clubs. As can be seen from the table, the cluster correlations between the various outcomes does not exceed 0·66. It thus seems that our results about the actual composition of the convergence clubs for the Penn World Table are not very reliable. The opposite is true for the Bernard and Durlauf (1995) data. As can be seen from Table V, the outcomes for these data are identical for $l = 3, 4, 5$. Hence, it seems that the increased sample size leads to increased robustness of the clustering results. In fact, our results for the data used by Bernard and Durlauf (1995) turn out to be very robust.

5.2. Composition of Convergence Clubs

The results for both the Penn World Table as well as the Bernard and Durlauf data have one striking feature in common. Namely, in both cases we find a relatively large number of convergence clubs. In particular, for the 112 countries in our sample taken from the Penn World Table we find 63 asymptotically perfect convergence clubs and 42 asymptotically relative convergence clubs.[3] For the 15 industrialized countries sampled in the Bernard and Durlauf data these numbers are 9 and 7 respectively.

Not only the number of convergence clubs that we find is surprising, even more is their composition, as can be seen from Tables BII–BIV in Appendix B. Although the sensitivity analysis in the previous section suggests the necessary caution in the interpretation of the convergence clubs found for the Penn World Table, some facts still seem to stand out. First, asymptotically perfect convergence seems to be more common for lower-income, especially African, economies, than for the industrialized countries. Table BII shows that the largest asymptotically perfect convergence clubs are all made up of low-income countries, most of them in Sub-Saharan Africa. The industrialized countries, apart from Finland and Iceland which are in a club with Trinidad Tobago, are all part of clubs of size two, for example {Belgium, Norway}, {Switzerland, United States}, {Denmark, Luxembourg}, {Germany, France} and {Great Britain, Netherlands}, or not converging to any other countries, like Australia, Austria, Canada, Ireland, Italy, Japan, New Zealand and Sweden. Second, we find almost no low-income countries that converge to high-income countries. The only exceptions are the asymptotically relative convergence clubs containing Kenya and Ecuador, i.e. clubs 1 and 3 in Table BIII. Finally, contrary to our expectations, we do not find much more asymptotically relative convergence than asymptotically perfect convergence.

For the industrialized countries in the Bernard and Durlauf data set, we again do not find much convergence. Furthermore, the convergence clubs we find do not seem to be determined by geographical location of the countries, nor do we obtain convergence for all the countries that are part of the European Union, as would be suggested by the results in Ben David (1993).

These results bring up two questions. First, what could be the possible reason for the lack of convergence obtained with our cluster algorithm? Second, what underlies this lack of convergence, i.e. is it caused by different deterministic or stochastic trends?

We can basically distinguish two types of possible reasons for our results. The first is purely statistical, while the second is economically theoretical.

One possible statistical reason might be that our method simply does not allow us to make reliable inference in the sample sizes of length 30 and 90 that we consider. That is, the tests introduced in Section 3 are very likely to reject the null hypothesis of convergence. Although $\bar{\varpi}_0$

[3] These are the results obtained using $p_{\min} = 0{\cdot}01$ and $l = 2$, as presented in Tables BII and BIII in Appendix B.

and $\bar{\varpi}_\mu$ might indeed be oversized in these small samples, the results that we presented were obtained with a 'significance level' of 1%, thus biasing our results in favour of convergence. Furthermore, using Johansen's (1991) cointegrating rank test, which tests for the null hypothesis of non-convergence, Bernard and Durlauf (1995) also find six common trends in output for their data set, suggesting the existence of, at least, six convergence clubs. Using a similar methodology, Daniel (1997) identifies two structural trends for per capita industrial production for the USA, the UK, and Japan. She shows that one of these trends can be identified by the price of oil. Again, she does not find convergence of the three economies that she considers. Another statistical reason for the lack of convergence we obtain is that the assumed data-generating process, i.e. equation (1) is possibly misspecified. In particular, the recent results by Ben-David *et al.* (1997) suggest that we might obtain more convergence when we allow for structural breaks.

At a theoretical level, it might be the case that convergence of productivity does not occur on the aggregate level of per capita income, but rather on sector-specific levels. Indeed, it seems to be likely that the knowledge spillover effects, as theoretically illustrated by, for example, Barro and Sala-i-Martin (1995), that are currently considered as a probable cause of convergence, are sector-specific. That is, if, for some reason, a technological innovation causes an increase in the productivity of steel production in the USA, spillover of this knowledge to steel producers in other countries will, most likely, cause faster growth of per capita income in countries with a relatively large steel manufacturing sector than in a country with little or no steel manufacturing. Bernard and Jones (1996) investigate convergence of productivity for various sectors of 14 industrialized economies. They indeed find a high degree of convergence on a sectoral level. It might thus be that there is convergence on sectoral levels, but that on an aggregate level economies do not converge due to the different compositions of their aggregate output.

If we want to answer the question whether deterministic or stochastic trends cause the observed lack of convergence, it is interesting to also consider convergence of growth rates. In order to see how much convergence of growth rates we find in our data, we performed an experiment in which we continued to cluster the asymptotically relatively converging clubs by testing for convergence of growth rates. For the Penn World Table data we ran into the problem that we obtained clusters that contain more than 30 countries, making any further inference impossible. For the Bernard and Durlauf (1995) data, however, we obtained two clubs with countries converging in growth rates. These clubs are $\{CAN, DEU, DNK, FIN, FRA, NLD, SWE, USA\}$, on the one hand, and $\{AUS, AUT, BEL, GBR, ITA, JAP, NOR\}$ on the other. These results suggest that convergence in growth rates is a much more widespread phenomenon than asymptotically relative convergence. This implies that, in the long run, the differences in log per capita productivity levels are mainly due to diverging stochastic rather than to deterministic trends.

6. CONCLUSION

In this paper we have tried to identify the number of convergence clubs and the countries which are included in these clubs. In order to do so, we have introduced a cluster algorithm that enabled us to endogenously select converging countries. The selection of converging countries is based on the necessary and sufficient conditions for two types of convergence: asymptotically perfect and asymptotically relative convergence. The first basically implies the convergence to identical log real GDP per capita levels, while the second implies convergence to constant relative real GDP per capita levels.

The empirical results suggest that convergence is not a very widespread phenomenon and is most common among low-income economies, especially in Sub-Saharan Africa. Furthermore, we find virtually no convergence of low-income economies to high-income economies. This lack of convergence seems to be mainly driven by diverging stochastic trends that underly international log real per capita GDP levels.

From an empirical point of view, two extensions of the above analysis would be particularly interesting. The first, following Daniel (1997), is the identification and possible explanation of the stochastic trends that drive international output. The second, following Ben-David *et al.* (1997), is repeating the above analysis, but allowing for structural breaks in the real per capita GDP series. From a theoretical point of view it would be useful to illustrate a scenario in which the sectors in, say two, multi-sector economies converge in productivity, but in which aggregate productivity diverges due to different shares in GDP for the various sectors.

APPENDIX A: PROOFS OF PROPOSITIONS

Proof of Proposition 1: We will only prove $\bar{\varpi}_0 = \bar{\varpi}_0^*$. The other result follows in a similar manner. We obtain that

$$\mathbf{S}_t^* = \sum_{s=1}^{t} \mathbf{x}_s^* = \mathbf{A} \sum_{s=1}^{t} \mathbf{x}_s = \mathbf{A}\mathbf{S}_t$$

Similarly, it can be easily seen that, $\hat{\mathbf{G}}_l^* = \mathbf{A}\hat{\mathbf{G}}_l\mathbf{A}'$. Thus

$$\begin{aligned} \bar{\varpi}_0^* &= T^{-2} \sum_{t=1}^{T} \mathbf{S}_t^{*\prime}[\hat{\mathbf{G}}_l^*]^{-1}\mathbf{S}_t^* \\ &= T^{-2} \sum_{t=1}^{T} \mathbf{S}_t'\mathbf{A}'[\mathbf{A}\hat{\mathbf{G}}_l\mathbf{A}']^{-1}\mathbf{A}\mathbf{S}_t \\ &= T^{-2} \sum_{t=1}^{T} \mathbf{S}_t'[\hat{\mathbf{G}}_l]^{-1}\mathbf{S}_t = \bar{\varpi}_0 \end{aligned}$$

■

Proof of Proposition 4: Without loss of generality, we will assume that

$$\mathbf{D} = \begin{bmatrix} \mathbf{0} \\ \underset{m\times m}{\tilde{\mathbf{D}}} \end{bmatrix}$$

where $\tilde{\mathbf{D}}$ is non-singular. This is without loss of generality because we can always find a non-singular transformation of the original vector process $\mathbf{x}_t$ for which this holds; Proposition 1 applies. We will first consider the asymptotic properties of the partial sums processes $\mathbf{S}_t$ and $\tilde{\mathbf{S}}_t$ and the asymptotic properties of $\hat{\mathbf{G}}_l$ separately. We will then combine them to obtain the consistency results claimed.

Consider the vector process $\mathbf{x}_t \in \mathbb{R}^k$ such that

$$\underset{k\times 1}{\mathbf{x}_t} = \underset{k\times 1}{\boldsymbol{\alpha}} + \underset{k\times 1}{\boldsymbol{\beta}}\, t + \underset{k\times m}{\mathbf{D}} \sum_{s=0}^{t-1} \underset{m\times 1}{\mathbf{v}_s} + \underset{k\times 1}{\mathbf{u}_t}$$

and let the ith elements of the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be denoted by α_i and β_i. Again without loss of generality, let $\mathbf{x}_t$ be such that there are non-negative integers k_1, k_2, k_3, k_4, k_5, possibly equal to zero, for which: $\alpha_i = \beta_i = 0$, for $0 < i \leqslant k_1$, $\alpha_i \neq 0$ and $\beta_i = 0$ for $k_1 < i \leqslant (k_1 + k_2)$, $\beta_i \neq 0$ for $(k_1 + k_2) < i \leqslant (k_1 + k_2 + k_3)$, where $k_1 + k_2 + k_3 = k - m$, $\beta_i = 0$, for $(k - m) < i \leqslant (k - m + k_4)$, and $\beta_i \neq 0$ for $(k - m + k_4) < i \leqslant (k - m + k_4 + k_5) = k$. Though their definition is complicated, these integers basically imply that the univariate processes in $\mathbf{x}_t$ are stacked such that the first k_1 are zero mean stationary, the next k_2 are level stationary, the next k_3 are trend stationary, the next k_4 contain a stochastic but no deterministic trend, and the last k_5 contain both a stochastic and a deterministic trend.

Applying the functional central limit theorem to $\boldsymbol{\eta}_t$ we obtain for $r \in [0, 1]$, that

$$T^{-1/2} \sum_{t=1}^{\langle Tr \rangle} \boldsymbol{\eta}_t \rightarrow \boldsymbol{\Lambda} \mathbf{W}(r)$$

where $\langle Tr \rangle$ is the integer part of Tr. Applying the continuous mapping theorem, it follows that

$$T^{-3/2} \sum_{t=1}^{\langle Tr \rangle} \sum_{\nu=0}^{t-1} \boldsymbol{\eta}_\nu \rightarrow \boldsymbol{\Lambda} \int_0^r \mathbf{W}(s)\, \mathrm{d}s$$

Let $\mathbf{W}_{k_1}(r)$ and $\mathbf{W}_{k_1+k_2}(r)$ be the first k_1 and $k_1 + k_2$ elements of $\mathbf{W}(r)$ respectively, and let $\tilde{\boldsymbol{\Lambda}}_{k_1}$ and $\tilde{\boldsymbol{\Lambda}}_{k_1+k_2}$ be square matrices consisting of the first k_1 and $k_1 + k_2$ rows and columns of $\boldsymbol{\Lambda}$ respectively. Similarly, let $\mathbf{W}_\nu(r)$ consist of rows $(k + 1)$ through $(k + m)$ of $\mathbf{W}(r)$ and let $\tilde{\boldsymbol{\Lambda}}_\nu$ consist of rows and columns $(k + 1)$ through $(k + m)$ of $\boldsymbol{\Lambda}$. Furthermore, let $\boldsymbol{\alpha}_{k_2}$ be the intercept vector associated with the level stationary processes in $\mathbf{x}_t$. Similarly, let $\boldsymbol{\beta}_{k_3}$ and $\boldsymbol{\beta}_{k_5}$ be the deterministic trend vectors associated with the trend stationary processes and the processes with both a deterministic and a stochastic trend in $\mathbf{x}_t$. Finally, let

$$\underset{k_4\times m}{\tilde{\mathbf{D}}_{k_4}}$$

be the rows of $\tilde{\mathbf{D}}$ associated with the processes in $\mathbf{x}_t$ with a stochastic but no deterministic trend.

We define the following matrices:

$$Z_T^0 = \begin{bmatrix} \mathbf{I}_{k_1} & & & & \\ & T^{-1/2}\mathbf{I}_{k_2} & & \mathbf{0} & \\ & & T^{-3/2}\mathbf{I}_{k_3} & & \\ & \mathbf{0} & & T^{-1}\mathbf{I}_{k_4} & \\ & & & & T^{-3/2}\mathbf{I}_{k_5} \end{bmatrix}$$

and

$$Z_T^{\mu} = \begin{bmatrix} \mathbf{I}_{k_1+k_2} & & & \\ & T^{-3/2}\mathbf{I}_{k_3} & & \mathbf{0} \\ & \mathbf{0} & T^{-1}\mathbf{I}_{k_4} & \\ & & & T^{-3/2}\mathbf{I}_{k_5} \end{bmatrix}$$

where $\mathbf{I}_k$ denotes the identity matrix of dimension k.

Using these matrices, we obtain for the asymptotic properties of the partial sums processes that

$$T^{-1/2} Z_T^0 \mathbf{S}_{\langle Tr \rangle} \to \begin{bmatrix} \tilde{\boldsymbol{\Lambda}}_{k_1} \mathbf{W}_{k_1}(r) \\ \boldsymbol{\alpha}_{k_2} r \\ \frac{1}{2}\boldsymbol{\beta}_{k_3} r(r-1) \\ \tilde{\mathbf{D}}_{k_4} \tilde{\boldsymbol{\Lambda}}_v \int_0^r \mathbf{W}_v(s)\,\mathrm{d}s \\ \frac{1}{2}\boldsymbol{\beta}_{k_5} r(r-1) \end{bmatrix} = \mathbf{Q}^0(r)$$

and, similarly, that

$$T^{-1/2} Z_T^{\mu} \tilde{\mathbf{S}}_{\langle Tr \rangle} \to \begin{bmatrix} \tilde{\boldsymbol{\Lambda}}_{k_1+k_2} \bar{\mathbf{W}}_{k_1+k_2}(r) \\ \boldsymbol{\alpha}_{k_2} r \\ \frac{1}{2}\boldsymbol{\beta}_{k_3} r(r-1) \\ \tilde{\mathbf{D}}_{k_4} \tilde{\boldsymbol{\Lambda}}_v \int_0^r \bar{\mathbf{W}}_v(s)\,\mathrm{d}s \\ \frac{1}{2}\boldsymbol{\beta}_{k_5} r(r-1) \end{bmatrix} = \mathbf{Q}^{\mu}(r)$$

where $\bar{\mathbf{W}}(r)$ is a demeaned $(k+m)$-dimensional standard Brownian motion, i.e.

$$\bar{\mathbf{W}}(r) = \mathbf{W}(r) - \int_0^1 \mathbf{W}(s)\,\mathrm{d}s$$

and $\bar{\mathbf{W}}_{k_1+k_2}(r)$ and $\bar{\mathbf{W}}_v(r)$ are defined similarly to $\mathbf{W}_{k_1+k_2}(r)$ and $\mathbf{W}_v(r)$.

For the asymptotic properties of $\hat{\mathbf{G}}_l$ we obtain the following. When we define

$$Z_T^G = \begin{bmatrix} \mathbf{I}_{k-m} & \mathbf{0} \\ \mathbf{0} & T^{-1/2} l^{-1} \mathbf{I}_m \end{bmatrix}$$

then

$$Z_T^G \hat{\mathbf{G}}_l Z_T^G \to \begin{bmatrix} \hat{\mathbf{G}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{D}} \hat{\boldsymbol{\Lambda}}_v [\int_0^1 \mathbf{W}_v^*(r) \mathbf{W}_v^*(r)'\,\mathrm{d}r] \tilde{\boldsymbol{\Lambda}}_v' \tilde{\mathbf{D}}' \end{bmatrix} = \mathbf{F}$$

where $\mathbf{W}^*(r)$ is a demeaned and detrended $(k+m)$-dimensional standard Brownian motion, i.e.

$$\mathbf{W}^*(r) = \mathbf{W}(r) + (6r-4)\int_0^1 \mathbf{W}(s)\,\mathrm{d}s + (-12r+6)\int_0^1 s\mathbf{W}(s)\,\mathrm{d}s$$

and $\mathbf{W}_v^*(r)$ consists of the last m rows of $\mathbf{W}^*(r)$. The matrix $\tilde{\mathbf{G}}$ consists of the first $(k-m)$ rows and columns of $\mathbf{G}$.

Combining the asymptotic results for the partial sum processes $\mathbf{S}_t$ and $\tilde{\mathbf{S}}_t$ and those for $\hat{\mathbf{G}}_l$, we thus obtain that

$$T^{-2}\sum_{t=1}^{T}\mathbf{S}_t' Z_T^0 (Z_T^G)^{-1}\hat{\mathbf{G}}_l^{-1}(Z_T^G)^{-1} Z_T^0 \mathbf{S}_t \to \int_0^1 \mathbf{Q}^0(r)'\mathbf{F}^{-1}\mathbf{Q}^0(r)\,\mathrm{d}r$$

and

$$T^{-2}\sum_{t=1}^{T}\mathbf{S}_t' Z_T^\mu (Z_T^G)^{-1}\hat{\mathbf{G}}_l^{-1}(Z_T^G)^{-1} Z_T^\mu \mathbf{S}_t \to \int_0^1 \mathbf{Q}^\mu(r)'\mathbf{F}^{-1}\mathbf{Q}^\mu(r)\,\mathrm{d}r$$

which implies the rates of consistency tabulated in Table AI. ■

Table AI. Rates of consistency

	Rate of consistency	
Case	$\bar{\varpi}_0$	$\bar{\varpi}_\mu$
$k_3 \neq 0$	T^3	T^3
$k_3 = 0, k_5 \neq 0$	T^2/l	T^2/l
$k_3 = k_5 = 0, k_4 \neq 0$	T/l	T/l
$k_3 = k_4 = k_5 = 0, k_2 \neq 0$	T	H_μ satisfied
$k_2 = k_3 = k_4 = k_5 = 0, k_1 \neq 0$	H_0 satisfied	H_μ satisfied

APPENDIX B: EMPIRICAL RESULTS

Table BI. List of countries

	Country		Country		Country
AGO	Angola	GIN	Guinea	NLD	Netherlands
ARG	Argentina	GMB	Gambia	NOR	Norway
AUS	Australia	GNB	Guinea-Bissau	NZL	New Zealand
AUT	Austria	GRC	Greece	PAK	Pakistan
BDI	Burundi	GTM	Guatemala	PAN	Panama
BEL	Belgium	GUY	Guyana	PER	Peru
BEN	Benin	HKG	Hong Kong	PHL	Phillipines
BGD	Bangladesh	HND	Honduras	PNG	Papua N. Guinea
BOL	Bolivia	HTI	Haiti	PRI	Puerto Rico
BRA	Brazil	HVO	Burkina Faso	PRT	Portugal
BRB	Barbados	IDN	Indonesia	PRY	Paraguay
BUR	Myanmar	IND	India	RWA	Rwanda
BWA	Botswana	IRL	Ireland	SEN	Senegal
CAF	Central African Rep.	IRN	Iran	SGP	Singapore
CAN	Canada	ISL	Iceland	SLV	El Salvador
CHE	Switzerland	ISR	Israel	SOM	Somalia
CHL	Chile	ITA	Italy	SUR	Suriname
CIV	Ivory Coast	JAM	Jamaica	SWE	Sweden
CMR	Cameroon	JOR	Jordan	SWZ	Swaziland
COG	Congo	JPN	Japan	SYC	Seychelles
COL	Colombia	KEN	Kenya	SYR	Syria
CPV	Cape Verde Is.	KOR	Korea	TCD	Tcad
CRI	Costa Rica	LKA	Sri Lanka	TGO	Togo
CSK	Czechoslovakia	LSO	Lesotho	THA	Thailand
CYP	Cyprus	LUX	Luxembourg	TTO	Trinidad/Tobago
DEU	West Germany	MAR	Morocco	TUN	Tunisia
DNK	Denmark	MDG	Madagascar	TUR	Turkey
DOM	Dominican Rep.	MEX	Mexico	UGA	Uganda
DZA	Algeria	MLI	Mali	URY	Uruguay
ECU	Ecuador	MLT	Malta	USA	United States
EGY	Egypt	MOZ	Mozambique	VEN	Venezuela
ESP	Spain	MRT	Mauritania	YUG	Former Yugoslavia
FIN	Finland	MUS	Mauritius	ZAF	South Africa
FJI	Fiji	MWI	Malawi	ZAR	Zaire
FRA	France	MYS	Malaysia	ZMB	Zambia
GAB	Gabon	NAM	Namibia	ZWE	Zimbabwe
GBR	United Kingdom	NER	Niger		
GHA	Ghana	NGA	Nigeria		

Table BII. Asymptotically perfectly convergence clubs (PWT)

No.	Countries					
1	BDI	HVO	MLI	MWI		
2	CAF	IND	NER	UGA		
3	GUY	JOR	SLV	SYC		
4	AGO	GHA	HTI			
5	BOL	LKA	PNG			
6	CIV	COG	MAR			
7	CPV	GNB	RWA			
8	FIN	ISL	TTO			
9	FJI	NAM	PER			
10	IRN	PRT	YUG			
11	MRT	PAK	SOM			
12	MYS	SWZ	TUR			
			Clusters with two countries			
13	BEL	NOR		14	BRA	SUR
15	CHE	USA		16	CHL	GAB
17	CMR	NGA		18	COL	JAM
19	CSK	GRC		20	CYP	SGP
21	DEU	FRA		22	DNK	LUX
23	DZA	GTM		24	EGY	ZWE
25	ESP	ISR		26	GBR	NLD
27	LSO	TGO		28	MDG	ZMB
29	MEX	URY		30	MOZ	SEN
31	MUS	ZAF		32	PAN	SYR
33	PRY	TUN		34	TCD	ZAR
			Twenty nine separate countries			
	ARG	AUS	AUT	BEN	BGD	BRB
	BUR	BWA	CAN	CRI	DOM	ECU
	GIN	GMB	HKG	HND	IDN	IRL
	ITA	JPN	KEN	KOR	MLT	NZL
	PHL	PRI	SWE	THA	VEN	

Listed according to size and alphabetical order.
$p_{\min} = 0{\cdot}01$ and $l = 2$.

Table BIII. Asymptotically relatively convergence clubs (PWT)

No.	Countries					
1	AUS	DNK	KEN	LUX	MUS	ZAF
2	AUT	ESP	ISR	ITA	PRI	
3	CAN	CSK	ECU	GRC	IRL	
4	BDI	HVO	MLI	MWI		
5	BGD	BUR	HND	NZL		
6	CAF	IND	NER	UGA		
7	GUY	JOR	SLV	SYC		
8	AGO	GHA	HTI			
9	BEN	GIN	VEN			
10	BOL	LKA	PNG			
11	BRB	IDN	THA			
12	CIV	COG	MAR			
13	CMR	CRI	NGA			
14	CPV	GNB	RWA			
15	FIN	ISL	TTO			
16	FJI	NAM	PER			
17	IRN	PRT	YUG			
18	MRT	PAK	SOM			
19	MYS	SWZ	TUR			
			Clusters with two countries			
20	ARG	GMB		21	BEL	NOR
22	BRA	SUR		23	BWA	MLT
24	CHE	USA		25	CHL	GAB
26	COL	JAM		27	CYP	SGP
28	DEU	FRA		29	DOM	SWE
30	DZA	GTM		31	EGY	ZWE
32	GBR	NLD		33	HKG	KOR
34	LSO	TGO		35	MDG	ZMB
36	MEX	URY		37	MOZ	SEN
38	PAN	SYR		39	PRY	TUN
40	TCD	ZAR				
			Two separate countries			
	JPN	PHL				

Listed according to size and alphabetical order.
$p_{\min} = 0{\cdot}01$ and $l = 2$.

Table BIV. Convergence clubs (Bernard and Durlauf-data)

	Asymptotically perfect			Asymptotically relative		
1	AUT	ITA	NOR	AUS	BEL	GBR
2	AUS	GBR		AUT	ITA	NOR
3	DEU	FIN		CAN	DNK	NLD
4	DNK	NLD		DEU	FIN	
5	FRA	SWE		FRA	SWE	
	Four separate countries			Two separate countries		
	BEL	CAN	JAP	JAP	USA	
	USA					

Listed according to size and alphabetical order.
$p_{\min} = 0{\cdot}01$ and $l = 4$.

ACKNOWLEDGEMENTS

We thank co-editor Steven Durlauf, two anonymous referees, and seminar participants at UC San Diego, Tinbergen Institute Rotterdam, and the 1998 North American Winter Meeting of the Econometric Society for their helpful comments. Bart Hobijn thanks the C. V. Starr Center at NYU for financial support.

REFERENCES

Barro, R. J. and X. Sala-i-Martin (1992), 'Convergence', *Journal of Political Economy*, **100**, 223–251.

Barro, R. J. and X. Sala-i-Martin (1995), 'Technological diffusion, convergence and growth', NBER working paper 5151.

Baumol, W. J. (1986), 'Productivity growth, convergence and welfare: what the long-run data show', *American Economic Review*, **76**, 1072–1085.

Baumol, W. J., R. R. Nelson and E. N. Wolff (eds) (1994), *Convergence of Productivity: Cross National Studies and Historical Evidence*, Oxford University Press, New York.

Ben-David, D. (1993), 'Equalizing exchange: trade liberalization and convergence', *Quarterly Journal of Economics*, **108**, 653–679.

Ben-David, D., R. L. Lumsdaine and D. H. Papell (1997), 'Unit roots, postwar slowdowns and long-run growth: evidence from *two* structural breaks', working paper.

Bernard, A. B. and S. N. Durlauf (1995), 'Convergence in international output', *Journal of Applied Econometrics*, **10**, 161–173.

Bernard, A. B. and S. N. Durlauf (1996), 'Interpreting tests for the convergence hypothesis', *Journal of Econometrics*, **91**, 161–173.

Bernard, A. B. and C. I. Jones (1996), 'Comparing apples and oranges: productivity convergence and measurement across industries and countries', *American Economic Review*, **86**, 1216–1238.

Bianchi, M. (1997), 'Testing for convergence: evidence from non-parametric multimodality tests', *Journal of Applied Econometrics*, **12**, 393–409.

Canova, F. and A. Marcet (1995), 'The poor stay poor: non-convergence across countries and regions', CEPR Discussion Paper 1265.

Caselli, F., G. Esquivel and F. Lefort (1996), 'Reopening the convergence debate: a new look at cross-country growth empirics', *Journal of Economic Growth*, **1**, 363–389.

Daniel, B. C. (1997), 'International interdependence of national growth rates: a structural trends analysis', *Journal of Monetary Economics*, **40**, 73–96.

Durlauf, S. N. and P. A. Johnson (1995), 'Multiple regimes and cross-country growth behavior', *Journal of Applied Econometrics*, **10**, 365–384.

Durlauf, S. N. and D. T. Quah (1998), 'The new empirics of economic growth', NBER Working Paper Series 6422. To appear in *The Handbook of Macroeconomics*.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, NJ.

Hobijn, B., P. H. Franses and M. Ooms (1998), 'Generalizations of the KPSS-test for stationarity'. *Econometric Institute Report 9802*, Erasmus University Rotterdam.

Islam, N. (1995), 'Growth empirics: a panel data approach', *Quarterly Journal of Economics*, **110**, 1127–1170.

Johansen, S. (1991), 'Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models', *Econometrica*, **55**, 251–276.

Jones, C. I. (1997), 'On the evolution of the world income distribution', *Journal of Economic Perspectives*, **11:3**, 19–36.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992), 'Testing the null hypothesis of stationarity against the alternative of a unit root', *Journal of Econometrics*, **54**, 159–178.

Mankiw, N. G., D. Romer and D. N. Weil (1992), 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics*, **107**, 407–437.

Newey, W. K. and K. D. West (1987), 'A simple, positive semi-definite, heteroskedasticity and auto-correlation consistent covariance matrix', *Econometrica*, **55**, 703–708.

Pritchett, L. (1997), 'Divergence, big time', *Journal of Economic Perspectives*, **11**(3), 3–17.

Quah, D. (1993), 'Galton's fallacy and tests for the convergence hypothesis', *Scandinavian Journal of Economics*, **95**, 427–443.

Quah, D. (1996a), 'Convergence empirics across economies with (some) capital mobility', *Journal of Economic Growth*, **1**, 95–124.

Quah, D. (1996b), 'Empirics for economic growth and convergence', *European Economic Review*, **40**, 1353–1375.

Summers, R. and A. Heston (1991), 'The Penn World Table (mark 5): an expanded set of international comparisons, 1950–1988', *Quarterly Journal of Economics*, **106**, 327–368.

